CLAIMS

1. A method of converting a voice signal (130) as spoken by a source speaker into a converted voice signal (150) the acoustic characteristics thereof resemble those of a target speaker, the method comprising:

    · a determination step (1) of determining a function for transforming acoustic characteristics of the source speaker into acoustic characteristics close to those of the target speaker on the basis of samples of the voices of the source and target speakers, and

    · a transformation step (2) of transforming acoustic characteristics of the source speaker voice signal (130) to be converted by applying said transformation function,

    the said method being characterized in that said determination step (1) comprises a step (1; 56) of determining a function for conjoint transformation of characteristics of the source speaker relating to the spectral envelope and of characteristics of the source speaker relating to the pitch and in that said transformation step (2) comprises applying said conjoint transformation function.

2. A method according to claim 1, characterized in that said step (1; 56) of determining a conjoint transformation function comprises:

    · a step (4X, 4Y) of analyzing source and target speaker voice samples grouped into frames to obtain for each frame information relating to the spectral envelope and to the pitch,

    · a step (16X, 16Y; 62X, 62Y) of concatenating information relating to the spectral envelope and information relating to the pitch for each of the source and target speakers,

    · a step (20; 70) of determining a model representing common acoustic characteristics of source speaker and target speaker voice samples, and

· a step (30; 80) of determining said conjoint transformation function from said model and the voice samples.

3. A method according to claim 2, characterized in that said steps (4X, 4Y) of analyzing the source and target speaker voice samples are adapted to produce said information relating to the spectral envelope in the form of cepstral coefficients.

4. A method according to claim 2 or claim 3, characterized in that said analysis steps (4X, 4Y) comprise respectively a step of achieving voice samples models as a summation of an harmonic, signal and noise, each achieving step comprising :

· a substep (8X, 8Y) of estimating the pitch of the voice samples,

· a substep (10X, 10Y) of synchronized analysis of the pitch of each frame, and

· a substep (12X, 12Y) of estimating spectral envelope parameters of each frame.

5. A method according to any one of claims 2 to 4, characterized in that said step (20; 70) of determining a model determines a Gaussian probability density mixture model.

6. A method according to claim 5, characterized in that said step (20; 70) of determining a model comprises:

· a substep (22, 72) of determining a model corresponding to a mixture of Gaussian probability densities, and

· a substep (24, 74) of estimating parameters of the mixture of Gaussian probability densities from an estimated maximum likelihood between the acoustic characteristics of the source and target speaker samples and the model.

7. A method according to any one of claims 2 to 6, characterized in that said step (1; 56) of determining at least one transformation function further includes a step (14X, 14Y; 60X, 60Y) of normalizing the pitch of the frames of source and target speaker samples relative to average values of the pitch of the analyzed source and target speaker samples.

8. A method according to any one of claims 2 to 7, characterized in that it includes a step (18; 50) of temporally aligning the acoustic characteristics of the source speaker with the acoustic characteristics of the target speaker, this step (18; 50) being executed before said step (20; 70) of determining a conjoint model.

9. A method according to any one of claims 1 to 8, characterized in that it includes a step (54) of separating voiced frames and non-voiced frames in the source speaker and target speaker voice samples, said step (56) of determining a conjoint transformation function of the characteristics relating to the spectral envelope and to the pitch being based only on said voiced frames and the method including a step (58) of determining a function for transformation of only the spectral envelope characteristics on the basis only of said non-voiced frames.

10. A method according to any one of claims 1 to 8, characterized in that said step (1) of determining at least one transformation function comprises only said step (1) of determining a conjoint transformation function.

11. A method according to any one of claims 1 to 10, characterized in that said step (1; 56) of determining a conjoint transformation function is achieved on the basis

of an estimate of the acoustic characteristics of the target speaker, the achievement of the acoustic characteristics of the source speaker being known.

5    12. A method according to claim 11, characterized in that said estimate is the conditional expectation of the acoustic characteristics of the target speaker the achievement of the acoustic characteristics of the source speaker being known.

10

13. A method according to any one of claims 1 to 12, characterized in that said step (2) of transforming acoustic characteristics of the voice signal (130) to be converted includes:

15    · a step (36) of analyzing said voice signal (130), grouped into frames, to obtain for each frame information relating to the spectral envelope and to the pitch,
    · a step (38) of formatting the acoustic information relating to the spectral envelope and to the pitch of the
20   voice signal to be converted, and
    · a step (40; 102) of transforming the formatted acoustic information of the voice signal (130) to be converted using said conjoint transformation function.

25   14. A method according to claim 9 in conjunction with claim 13, characterized in that it includes a step (100) of separating voiced frames and non-voiced frames in said voice signal (130) to be converted, said transformation step comprising:

30    · a substep (104) of applying said conjoint transformation function only to voiced frames of said signal (130) to be converted, and
    · a substep (106) of applying said transformation function of the spectral envelope characteristics only to
35   non-voiced frames of said signal (130) to be converted.

15. A method according to claim 10 in conjunction with claim 13, characterized in that said transformation step comprises applying said conjoint transformation function to the acoustic characteristics of all the frames of said

5    voice signal (130) to be converted.

16. A method according to any one of claims 1 to 15, characterized in that it further includes a step (44; 110) of synthesizing a converted voice signal (150) from

10   said transformed acoustic information.

17. A system for converting a voice signal (130) as spoken by a source speaker into a converted voice signal (150) the acoustic characteristics thereof resemble ones

15   of a target speaker, the system comprising:
     · means (124) for determining at least one function for transforming acoustic characteristics of the source speaker into acoustic characteristics similar to ones of the target speaker on the basis of voice samples as

20   spoken by the source and target speakers, and
     · means (136, 138) for transforming acoustic characteristics of the source speaker voice signal (130) to be converted by applying said transformation function,
     the said system is characterized in that said means

25   (124) for determining at least one transformation function comprise a unit (126) for determining a function for conjoint transformation of characteristics of the source speaker relating to the spectral envelope and of characteristics of the source speaker relating to the

30   pitch and in that said transformation means include (136) for applying said conjoint transformation function.

18. A system according to claim 17, characterized in that it further includes:

35   · means (132) for analyzing the voice signal (130) to be converted, adapted to produce information relating

to the spectral envelope and to the pitch of the voice signal (130) to be converted, and

· synthesizer means (140) for forming a converted voice signal from at least said spectral envelope and pitch information transformed simultaneously.

19. A system according to claim 17 or claim 18, characterized in that said means (124) for determining at least one acoustic characteristic transformation function further include a unit (128) for determining a transformation function for the spectral envelope of non-voiced frames, said unit (126) for determining the conjoint transformation function being adapted to determine the conjoint transformation function only for voiced frames.